

DOCUMENT RESUME

ED 093 912

TM 003 728

AUTHOR Ford, Valeria A.
TITLE Everything You Wanted To Know About Test-Wiseness.
PUB DATE Aug 73
NOTE 30p.
EDRS PRICE MF-\$0.75 HC-\$1.85 PLUS POSTAGE
DESCRIPTORS Guessing (Tests); *Literature Reviews; Multiple Choice Tests; *Response Style (Tests); Testing Problems; *Test Wiseness

ABSTRACT

The purpose of this paper is to acquaint the reader with the topic of test-wisness. The first section of this paper presents a series of multiple-choice items. The reader is asked to respond to them and is encouraged to read carefully the remainder of this paper for techniques which could improve test-taking performance. The next section defines and elaborates on the term test-wisness as it relates to other psychometric terminology. Then, an attempt is made to distinguish test-wisness from several other concepts frequently used indiscriminately. The third section, entitled "Selected Aspects of Test-Wisness," outlines a set of behaviors that a test-wise examinee should employ as he takes a test. Empirical research to support the content of the outline is elaborated in detail in the succeeding section. Finally, a rationale for the use of a test-wisness learning program to be used in the school is presented, and several possible procedures as to how such a program could be implemented are suggested. (Author/BB)

TN. 003 728

ED 093912

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

EVERYTHING YOU WANTED
TO KNOW ABOUT
TEST-WISENESS

Valeria A. Ford
Summer Graduate Student, 1973
Educational Testing Service
August 9, 1973

An Overview

The purpose of this paper is to acquaint the reader to the topic of test-wiseness. Test-wiseness, as an issue in educational and psychological measurement, has been investigated rather extensively and independently by numerous researchers, however, no attempt, familiar to this writer, has been made to organize the research into a systematic presentation. In acquainting the reader to the topic, it is hoped that an understanding of one of the important influences effecting test performance will be acquired.

The first section of this paper presents a series of multiple-choice items. The reader is asked to respond to them and is encouraged to read carefully the remainder of this paper for techniques which could improve test-taking performance.

The next section defines and elaborates upon the term test-wiseness as it relates to other psychometric terminology. Then, an attempt is made to distinguish test-wiseness from several other concepts frequently used indiscriminately.

The third section, entitled "Selected Aspects of Test-Wiseness" outlines a set of behaviors that a test-wise examinee should employ as he takes a test. Empirical research to support the content of the outline is elaborated in detail in the succeeding section.

Finally, a rationale for the use of a test-wiseness learning program to be used in the school is presented and several possible procedures as to how such a program could be implemented are suggested.

Your Test-Wiseness Quotient

This is a test of test "wiseness." Answer each question after you have decided on the response you feel will earn you the highest score. Each question can be answered by employing several test-wiseness techniques.

1. The typical "potassium potential" across a nerve-axon membrane is
 - A. -90 m V
 - B. -125 m V
 - C. -35 m V
 - D. -60 m V
2. The Municipal Corporation Bills, enacted by Parliament in 1835
 - A. Guaranteed the rights and freedoms of large urban industries
 - B. eliminated "rotten boroughs" by redistricting urban areas
 - C. provided a plan of government for most areas except London and some small towns
 - D. gave Parliament the right to interfere with private industry
3. The Fisher Act, legislated by Parliament
 - A. established a national system of public education with tax-supported, public elementary and secondary schools
 - B. reformed "rotten boroughs"
 - C. reduced the power of the House of Lords
 - D. created a system of socialized medicine
4. Inhibition of the motor neurons may be brought about by
 - A. an anodic increase in the Nernst potential
 - B. elicitation of a synaptic potential across the postsynaptic membrane
 - C. blockage of acetylcholine at the synaptic junction by chemical inactivation
 - D. intramuscular administration of physiological saline solution
5. "The Salad" is a fairy tale by:
 - A. Percy Bysshe Shelly
 - B. the Brothers Grimm
 - C. George Meredith
 - D. John Donne
6. In his first message to Congress, President Theodore Roosevelt:
 - A. demanded greater control of large corporations by the federal government
 - B. avoided the tariff question as politically too controversial
 - C. called for the reaffirmation of an economic policy with strongly favored free enterprise
 - D. vehemently advocated a higher tax schedule on imports
7. The Compromise between Democrats and Republicans after the election of 1876 resulted in
 - A. federal aid to Southern railroads
 - B. a higher federal tariff to assist the automobile industry
 - C. the entrance of Texas into the Union as a slave state
 - D. a treaty with Joseph Stalin
 - E. the Fugitive Slave Law

8. In the autonomic division
 - A. the cell bodies of the motor neurons are located exclusively within the brain and spinal cord
 - B. the cell bodies of the motor neurons may be found near the organs supplied
 - C. the fibers cannot inhibit the organ which is supplied
 - D. the organs supplied may be inhibited as well as stimulated by the fibers
9. The population of Frankton is more than
 - A. 50 million
 - B. 60 million
 - C. 70 million
 - D. 80 million
10. When Beston crystals are added to water
 - A. heat is given off
 - B. the temperature of the solution rises
 - C. the solution turns blue
 - D. the container becomes warmer

Do not be discouraged by these items. By reading the remainder of this paper, you will gain a greater understanding of the topic of test-wisness and, hopefully your sophistication in test-taking techniques will be improved.

TEST-WISENESS: THE STATE OF THE ART

Test-wisness and test score variance

The level of attainment of a group or an individual can be measured by a variety of objective techniques -- the best known among them being standardized tests. Most of the existing objective tests are composed of items which sample a learner's behavior at the precise moment of testing. The score on these tests is the outcome of the learner's total experience, which in turn is a result of (1) a combination of such factors as his overall innate ability, (2) his general environment, and (3) his special knowledge in the subject matter tapped by the test. In addition, certain studies have noted the influence of "test-wisness" on test scores (Dunn & Goldstein, 1959; Millman & Setijadi, 1966; Vernon, 1962).

Test-wiseness has been defined as the ability to manifest test-taking skills which utilize the characteristics and formats of a test and/or test-taking situation in order to receive a score commensurate with the abilities being measured (Oakland & Wehlert, 1971). Other names for this trait have been offered such as "test sophistication" (Erickson, 1972) or "test-wisdom" (Preston, 1964).

If it could be assumed that all but inborn factors were equal for all learner's (e.g., environment and/or training), the test scores obtained could be considered as representing the learner's true ability. If the nature of the test material is such that some students are at an advantage from having past experiences in test-taking, then some inaccuracy in measurement can occur.

A test-wise examinee can be expected to obtain a higher score on an aptitude, achievement, or teacher-made test, than an equally competent examinee who lacks test sophistication. Minority spokesmen argue that part of the reason for lower scores for their children results from a lack of familiarity with the "tricks" of taking a test. If, in fact, test scores for children of certain backgrounds are lower as a result of a lack of familiarity with some non-content factor (e.g., not knowing when to guess, misunderstanding instructions), then the argument may be valid.

Test constructors have attempted to reduce the variance in test scores resulting from degrees of test sophistication by writing clearer directions, eliminating specific determiners (e.g., often, seldom, always, must, etc.); however, various authors (Ebel, 1965) have conjectured that test-wiseness produces an additional source of test-related influence in educational test scores beyond that due to item content or random error. In an effort to

assure that knowledgeable students are not unduly penalized by poor examination skills, considerations have been given to training individuals who lack test sophistication. Woodley (1972) considers test-wiseness to be a cognitive factor, one which is measurable and subject to change either through specific test experience or training in a test-taking strategy. Consequently, rather than viewing test-wiseness as insignificant or undesirable, a growing number of recent investigations, suggests that people should be given training in "how to take tests."

Test-wiseness and subject matter coaching

Several terms have been used to express the many kinds of short-term intervention programs which have been designed to effect score gains of students taking tests intended to measure school achievement and general aptitude. Numerous research efforts across a whole range of approaches have attempted to cause some changes in short periods of time. Differences in the kind of training and/or content of instruction incorporated in the intervention program or experimental treatment constitute a workable distinction between the terms. In order to facilitate discussion for the remainder of this paper, the two terms will be differentiated.

Several studies cited in the research literature have focused on substantive test content and have come to be classified under the generic of "coaching studies;" intensive tutoring by a commercial tutoring school in vocabulary, reading, concept mastery, and problem-solving (Whitla, 1962); accelerated reading (Coffman & Parry, 1967); and commercial coaching with questions similar to those on SAT (Frankel, 1960). Findings by those who have explored the effects of coaching intervention on test scores (Dyer, 1953; French, 1955; French and Dear, 1959) have, historically,

yielded negligible results. Anastasi (1968) makes the point that "the more closely instruction is restricted to specific test content, the less likely is improvement to extend to criterion performance."

In what is commonly known as the Tennessee Study, Roberts & Oppenheim (1966) developed a specially prepared linear programmed material, designed to foster an analytic approach to the tasks required on college admissions in the verbal & mathematical areas. Instruction in either of the two areas was given to 11th grade students. Alternate forms of the PSAT, administered as pre- and post-tests were used to evaluate the effectiveness of the special instruction. While there were statistically significant differences in gain scores between some of the experimental groups, and their controls, the outcome of the study, like the outcome of earlier studies, was generally negative. The authors concluded that "the magnitude of the gains observed in the experimental groups were so small that it does not seem reasonable to expect that similar short term instruction given on a wide scale would be of significant benefit." The authors caution that the seven and one-half hour of instruction is less than what has been given in other coaching studies or at commercial coaching schools.

Based upon several coaching studies investigated by the College Entrance Examination Board, the trustees of the CEEB issued a detailed statement asserting that the general score gains directly attributable to coaching generally amount to fewer than 10 points, "a difference of such small magnitude on the SAT score scale of 200 to 800 that it is unreasonable to expect it to affect admissions decisions (College Board, 1965, p. 4)."

Recently, an investigation of the effects of special coaching on mathematics items on GRE test performance (Pike & Evans, 1972) has met with some success, however, the method of instruction has been criticized (Flaughner, 1972) because of the excessive length and intensity of the program.

In contrast to coaching studies, test-wiseness studies have focused upon instruction in general test-taking strategies (Moore, 1968), instruction in strategies specific to items grouped by item format (Moore, 1971), practice with the kinds of tests the examinee will encounter in the testing situation (Droege, 1966) and motivational activities prior to testing (Omwig, 1971). Test-wiseness is generally independent of the examinees knowledge of the trait or the subject matter for which the items are supposedly measures. Thus, in test-wiseness studies, instruction in subject matter material has been avoided. The primary concern has been to concentrate on the cognitive skills prerequisite to successful test taking. Since 1965, a variety of learning experiences designed to facilitate test-wiseness or specific skills related to test-wiseness have been developed. In most cases, the studies have indicated that the levels of test-wiseness can be increased through instruction (Moore, Schutz & Baker, 1966; Wahlstrom & Boerma, 1968; Oakland & Weilart, 1971).

Effects of Practice on Test

Familiarity with the format used in standardized tests pertain to a person's test wiseness. Millman and Setijadi's (1966) study demonstrated the disadvantage under which students operate when taking a test with an unfamiliar format. Comparisons were made of the performance of Indonesian and American students on tests using open-ended and multiple-choice questions involving arithmetic computation and vocabulary. The Indonesian students did relatively less well on the multiple-choice items than they did on the open-ended questions. This differential performance presumably was due to their having had no prior experience with multiple-choice items.

It has often been assumed that the experience of taking aptitude tests often leads to an increase in test scores upon retesting. Systematic study of the effect of practice supports this assumption.

Kreit (1967) examined the effects of test practice on the acquisition of test-taking skills of third grade pupils. Four different group intelligence tests were administered to the experienced group, while the control group received only pre- and post-tests. Increases in test-taking skills were acquired by the experimental group, presumably from increased exposure to a variety of tests.

Droege (1966) investigated the long-range effects of practice on the General Aptitude Test Battery (GATB), with an alternate form after 1 year, 2 years, and 3 years. The major findings revealed that increases in aptitude means of scores between initial testing and retesting show that all of the increases were significant, and that the effects of practice appear to be operating for all subsamples even when the interval between testings is as long as 3 years.

Familiarization with Item Types

In a study previously cited (Roberts & Oppenheim, 1966), materials consisting of "courses" of programmed instruction in five areas (four devoted to specific item types and a fifth to test-taking in general) were developed. Programs for the various item types (sentence completion, antonym, analogies, reading comprehension) and to create an analytic approach to test questions. Practice questions with the various item types were provided. Statistical significance was reported for post test scores of experimental and control groups, however, this difference was attributable to a substantial drop in the post-test scores of the control group.

Investigations into the various aspects of test-wisemess were basically a result of a descriptive paper written by Millman, Bishop and Ebel (1965), which provided a conceptual framework for the topic. Based upon empirical studies from the last fifteen years, an attempt is made to identify useful strategies for taking tests which coincide with/or expand the informal "test-wisemess principles" outlined by Millman et al. The following outline was developed from a variety of learning programs related to test-wisemess.

Selected Aspects of Test Wisemess

I. Instructions in general test-taking strategies

A. Follow directions

1. listen carefully over an expanded period of time [standardized tests contain important directions that are read orally.]
2. pay close attention to directions concerning allotted time and penalty contingences.
3. be sure to understand directions [if directions are not understood, ask the examiner immediately for clarification].

B. Properly mark responses (item response mode)

1. understand the proper way to mark responses [if the appropriate way to mark responses is not clear, immediately ask the examiner].
2. remember that different parts of standardized tests require the ability to make appropriate responses, as required, for each of the different parts of the test.
3. remember to select the correct alternative, identify the alternatives' code and item number, transfer this information from the test booklet (if necessary) to the answer sheet.

C. Use time efficiently

1. first answer questions of which you are sure.
2. then do the more difficult and time-consuming questions.
3. check answers during any remaining time to assess correctness and avoid careless mistakes.
4. if time permits, return to those questions previously omitted.
5. don't ponder over alternatives.

D. Learn when and when not to guess

1. guess only after an honest attempt has been made to answer the question.
2. guess, especially if one or more wrong answers can be identified.
3. if there is no penalty for guessing, answer every question since any question answered correctly will add to the total number correct.
4. if the test is a power test, with a penalty for guessing, omit any question for which at least one choice cannot be eliminated.

E. Miscellaneous tips

1. do not hesitate to change an answer, if you feel you should, since the percentage of wrong to right revisions tends to exceed right to wrong.
2. examine carefully all possible responses before attempting to choose the correct answer.
3. learn to concentrate attention to test-relevant variable while directing attention away from self-evaluative ruminations.
 - a. don't worry about your performance or how well others are doing.
 - b. don't ruminate over alternatives.

II. Instructions in strategies specific to item format

A. Answering analogy questions

1. learn to recognize common kinds of relationships presented in analogy questions [purpose, cause effect, part-to-whole, part-to-part, action to object, object-to-action, synonym, antonym, degree, characteristic, sequence, grammatical, numerical, association.]
2. plan a strategy for answering analogy questions - learn to form an immediate associate relatedness among the words in an analogy item.

B. Answering multiple-choice questions

1. eliminate options known to be incorrect and choose from the remaining alternatives.
2. eliminate similar options i.e., options which imply the correctness of each other.
3. eliminate those options which include specific determiners.
4. select the option which resembles an aspect of the stem, such that a name or phrase is repeated in both the stem and alternative.
5. look for correct alternatives which are longer and in some cases more precise and specific in meaning than other alternatives.
6. be alert to grammatical clues in the stem.
7. watch for overlapping distractors in which the truth of one alternatives implies the correctness of several others.
8. beware of one item "giving away" the answer to another item occurring in a latter part of the test.

As stated earlier, the preceding outline is based upon recent findings in test wiseness. Research literature supporting the content of the outline will be summarized.

A REVIEW OF THE LITERATURE

Research in Support of Changing Answers

Often students are heard to ask whether or not they should reconsider and possibly change their response to a test item. Writers on the topic of test-taking are, for the most part, in agreement concerning the advisability of answer changing. Millman et. al. suggests that the tendency to evaluate and judiciously change one's item response is a basic aspect of test-wiseness. Conclusions from recent research (Bath, 1967; Reile and Briggs, 1952; Stanley, 1971; Copeland, 1972) suggest that it is apparently advisable to change one's responses, since the typical result is an improvement in test scores.

In order to ascertain the effect on test grades of changing answers, Lynch and Smith (1972) administered a multiple choice test to 178 undergraduate students. Of the 294 response changes they identified, 68% were from wrong to right as compared with 32% for the reverse. This finding underscores earlier studies (Berrein, 1939; Lowe and Crawford, 1939), however, the authors reported that repeated changes on the same item did not improve a person's score.

Research on Guessing Strategies

When students are told they will be given a multiple-choice test, they sometimes ask if they will be penalized for guessing. Most often students are advised not to guess on standardized multiple-choice tests. Earlier evidence (Sheriffs and Boomer, 1954) has indicated that when examinees are requested

to respond to items previously omitted, their mean test score will increase. However, recent experiments investigating the effect of guessing on test scores report somewhat conflicting findings.

Taylor (1966) studied three different instruction groups ("do not answer unless certain;" "do as well as you can;" "answer all questions") and found no differences in mathematics scores due to instructions. On the other hand, Hammerton (1965) found that when examinees were instructed to answer all questions, they scored higher on vocabulary tests than when instructed to leave out questions if not confident of the answer.

The finding of Michael et. al. (1963) indicated that the average level of performance increased when guessing was encouraged and decreased with instructions advising against guessing. Another researcher (Ebel, 1968) concludes from his study with four separate, true-false tests that "students seeking highest scores on a test are well advised to answer all questions even when the usual correction is applied (their blind guesses to true-false items tend to be correct more than half of the time.)" Slakter (1968) sought to determine which of the following strategies would result in a higher test score: (a) you will receive one point for each correct answer, lose $1/3$ of a point for each incorrect answer, and neither receive nor lose points for questions you omit: usual guessing strategy (b) answer every question, whether you are sure of the answer or not: high guessing strategy. With these conventional directions, the results demonstrate that the "high guessing" group averaged five points higher than the "usual guessing" group, which indicated that the guessing effect was significant.

With the exception of the Taylor study, which has been criticized on methodological bases (Slakter, 1968), it appears that an examinee would tend to increase his test score by answering all items, even though a penalty for incorrect responses may be in effect.

In a somewhat different approach to research on guessing, Moore, Shultz and Baker (1966) developed a programmed text designed to help students develop optimal guessing strategies for responding to both speed and power tests. Several sample test directions gave the learner experience in identifying the important test contingencies. Various item sets provide guided practice in utilizing efficient procedures with respect to "guessing" and pacing of responses.

<u>Penalty for Guessing</u>	<u>Power Test</u>	<u>Speed Test</u>	<u>High Guessing</u>	<u>Low Guessing</u>
Yes	Yes	No	No	Yes
No	Yes	No	Yes	No
No	No	Yes	Yes	No
Yes	No	Yes	No	Yes

Table 1 reviews penalty and time characteristics, in addition to the optimal guess strategy under each condition. The data for the study consisted of guessing scores on four criterion tests in which directions regarding penalty and time contingencies varied (e.g. power test with no penalty for guessing, speed test with penalty for guessing). Research with the text showed a statistically significant difference between experimental and control groups in terms of guessing performance.

In another research study addressing this problem, Callenbach (1973) provided instruction to students regarding guessing. Instruction focusing on guessing stressed that students should (1) first answer questions to which one is sure and (2) guess only after an honest attempt has been made to answer the questions. Although these guessing instructions were only a part of the total instructional program in content - independent test-taking techniques, the overall results revealed a significant difference between experimental and control groups on the Stanford Reading Test.

Research Related to Solving Analogy Problems

Analogy items of the form accompanying the Miller Analogy Test (MAT):

light: dark: pleasure:

a. picnic b. day c. pain d. night

are an integral part of the testing many levels: high school, college, and graduate school. Some research suggests the possibility that the task of completing analogy questions, in many instances, is unique to many examinees, and that having exposure to such problems tends to facilitate performance. Colardarci (1960) reported that MAT retest performance was substantially higher than performance on an initial test. Moore (1966) correlated an inverse relationship between the time needed to complete sample questions and MAT scores. He suggests that not understanding the analogy task might pose a problem for many students. In a later study (Moore, 1971) the analogy test item was viewed as having certain characteristics that could be systematically taught. He hypothesized that examinees who were taught about the analogy item would be more able to deal with the relationships implied by the subject matter. Materials consisted of instruction on two aspects of test-wiseness in an analogy situation: question format, kinds of analogies, and use of the colons. The analysis showed a significant effect for the treatment factor. The findings of this study, together with others in the area of test-wiseness, suggests the possibility of improving the performance of examinees by brief instruction.

How to Approach Multiple-Choice Items

Not only have analogy items been considered for use in instructional programs in test-wiseness, considerable attention has been devoted to multiple-choice questions. Slakter, Koehler, and Hampton (1970) developed a programmed text to teach certain

selected aspects of test wiseness to 5th through 11th grades: (1) stem options, (2) absurd-options, (3) similar options, and (4) specific determiners. The following are examples of the kinds of items used to instruct students in these four basic types of clues.

1. stem-options

The revolution of the Four Hundred took place

- (1) after the unfortunate and long-lamented death of Theramenes
- (2) after the Sicilian expedition by the four hundred swordsmen of Alcibiades.
- (3) before the desperately sought peace of Nicias.
- (4) after the death of Alcibiades.

2. Absurd-options

The state with the highest total population in the United States in 1950 was

- (1) Chicago
- (2) Canada
- (3) New York
- (4) Washington, D.C.

3. Similar options

The line that is normal to the curve $y = x^2 + 2x - 3$ at the point (1,0) intersects the curve at one other point. What is the abscissa of this other point?

- (1) -3.25
- (2) -18/4
- (3) -4.50
- (4) -4 1/2

4. Specific determiners

French diplomacy between 1789 and 1815 was

- (1) completely unsuccessful in establishing better trade relations
- (2) to find some success in spreading revolutionary ideas to the rest of Europe
- (3) totally committed to the preservation of the monarchical form of government in Europe
- (4) absolutely successful in winning sympathy for Napoleon

Wahlstrom and Boersma (1968) conclude that the teaching of test-wise techniques led to improved performance on multiple-choice tests among the ninth graders they studied. Students were assigned to experimental, control, and placebo groups. The treatments were, respectively: instruction in principles of test-wiseness (a modified version of the principles of Millman et. al., 1966); watching television; and discussion of occupational information. Each group received pre- and post-tests in social studies which employed either "good" or "poor" multiple-choice items in an attempt to evaluate the effects of test-wiseness in objective examinations. Differences between pre- and post-test mean scores for control and placebo groups were all non-significant, however, the experimental group scored significantly higher than its control group.

In a study similar in design and methodology to Slaker et. al. (1970), a program designed specifically for use by adults was aimed at improving the general test-taking skills of candidates for the professional designation of Chartered Life Underwriters (Woodley, 1973). In order to determine the actual level of test wiseness in this population, a test was designed to measure whether or not the examinee could arrive at the appropriate answer to an item by recognizing and eliminating (1) similar options; and (2) absurd options, and (3) selecting an option which has a logical relationship with the stem.

Instructional packages, one for each of the test wiseness strategies taught, was used as the intervention for the experimental group. The test wiseness scale described earlier served as the pre- and post-test (internal criterion). All candidates, also took the Chartered Life Underwriter examination (external criterion). The findings showed that training in test wiseness produced the desired effect on the internal criterion, but did not on the external criterion.

Gibb (1964) was also concerned with cues found in multiple choice items and characterized test-wiseness as a "secondary cue response." He administered a preliminary assessment of test-wiseness to a group of _____ graders, which included seven types of secondary cues:

- (1) sound, word or phrase repetitions between stem and correct alternative
- (2) incorrect alternatives that are grossly unrelated or bear an absurd relationship to the stem
- (3) inclusion of words such as "all," "none," "nerves," etc. in the phrasing of incorrect alternatives
- (4) correct alternatives that are clearly more precise in meaning, more fully qualifying, or more specific than the incorrect alternatives
- (5) correct alternatives that are visibly longer than incorrect alternatives
- (6) grammatical cues, such as differences in number or tense of verbs, nouns, or the indefinite article between stem and alternatives
- (7) one item "giving away" the answer to another item occurring in a different part of the test

An experimental group was given brief training in the nature and recognition of these secondary cues. Following training, the experimental group and its control were given a 70 item, multiple-choice instrument of test-wiseness. Results showed that the trained group profited markedly from its control, and its performance was significantly higher. He also indicated that performance on his items was not determined by knowledge of the subject matter.

Based upon a survey of achievement batteries commonly used in the primary grades and from a careful analysis of the Stanford Reading Test, primary I, Callenbach (1972) constructed experimental materials for instruction and practice in (1) following orally administered directions, (2) response marking, (3) using time efficiently, and (4) guessing. Instructions stressed that: (1) standardized tests contain important directions that are read orally; (2) in order to answer questions correctly, the student must listen intently over an extended period of time; (3) in order to

answer the questions correctly, students must understand the directions; (4) different parts of standardized tests require the ability to make appropriate responses, as required for each of the different parts of the test; (5) when the student does not understand the directions, or the proper way to mark responses, he should ask for clarification; (6) students should first answer questions he knows answers to; (7) difficult and time consuming questions should be answered last; (8) answers should be checked during any remaining time to assess correctness and to avoid careless mistakes.

The standardized reading test scores of test-naïve students can be significantly raised through instruction and practice in content-independent test-taking techniques.

Training Students to Be Better Test Takers: Motivational Aspects

Omwig (1971) investigated the effects of motivational and teaching activities carried out prior to testing on test performance. Experimental students were given individual counseling on their past record of standardized achievement testing. A brief discussion on test taking included (a) why the test was administered, (b) the importance of following test directions, (c) marking the answer sheet correctly, (d) determining possible correct responses on multiple choice items before marking the desired response, (e) factors involved in scoring the test, and (f) an explanation of test norms and percentile ranks. When test results (Iowa Tests of Educational Development) for the experimental and control groups were compared, the experimental group achieved higher scores on all subtests. However, only subtests dealing with proficiency in English and mathematics were significantly higher. Although the results are not overwhelmingly conclusive, the author contends that improved test results are obtainable when students become personally involved in, motivated by, and interested in the testing program.

Test-Wiseness With Young Children

Most of the previous studies on test-wiseness have dealt with learners in upper elementary grades or above. Oakland (1972) attempted to examine certain variables which apply to test-wiseness factors in assessing preschool and primary-grade children. His intent was to familiarize children to the format and language used in standardized tests. The initial phase of the study identified specific abilities which appeared to be prerequisites to taking various standardized readiness tests (e.g. Metropolitan Readiness Test, Lee-Clark Reading Readiness Test). Having identified these abilities, curricular materials were designed to facilitate the development of these abilities.

The test-wiseness materials utilized several different formats in order to provide children with an opportunity to learn to respond to a variety of different tasks. No attempt was made to increase content area knowledge. The materials do the following:

1. Begin with few items and options per page and gradually increase them in number until the page is similar in appearance to an actual test page.
2. Provide practice in working in columns and rows.
3. Teach the concepts right, left, up, down, opposite, most like, same, and different.
4. Gradually increase the number of options per frame from two through five.
5. Encourage children to examine carefully all possible responses before choosing the correct one.
6. Progress from big pictures and words with few on a page to small pictures and words with several on a page, again until the final page is similar to an actual test page.
7. Provide practice in putting a mark directly on, directly under, or in the circle under the correct response.

8. Teach the children that biggest can mean the most and that a pencil may be called a marker.
9. Use both dotted lines and heavy black lines to separate the criterion from the options.
10. Refer to each page as a test in order to get children accustomed to the word.
11. Encourage children to ask questions if they do not understand the test directions.
12. Teach children to use a marker for keeping their places as they progress on each page.
13. Gradually increase the length of time children are encouraged to remain task oriented.

The Metropolitan Readiness Test was used to evaluate the effectiveness of the materials. Results revealed significant group differences between the experimental and control groups for total scores, and the "Matching" sub-test.

In addition to the various studies that have been reviewed in this paper, other attempts have been made to investigate aspects of test-wiseness in several different approaches. Several of these attempts will be described below.

Non-experimental Attempts at Test-Wiseness Programs

Inasmuch as disadvantaged persons tend to have fairly restricted experience in taking employment and similar aptitude, intelligence and job knowledge tests, it was reasoned that the lack of such exposure leaves them without test-taking skills. The U. S. Department of Labor Manpower experimental and demonstration projects have explored several approaches in the pretesting preparation of out-of-school and out-of-work youth and adults. To develop positive attitudes toward testing and to build test-taking confidence, Seiler (1971) reports that Labor Departments projects found the following of value:

1. fully explain why tests are given and how they can help the individual taking them.

2. allow the applicant to have a look at sample tests and take practice tests without time limits
3. answer questions about tests and explain errors made on practice tests
4. expose applicants to non-threatening test settings (test center/room environment)
5. permit applicants to test each other to become familiar with tests and to understand better why they are required

Seiler also describes an E & D project in Tennessee that found it useful to give instructions on how to choose answers; when to guess on true-false, multiple-choice, and matching type tests; the advisability of answering the easier questions first; and the benefit of using the process of elimination to find correct answers. Although Seiler presents no evidence on the effectiveness of these projects, he maintains that pretesting orientation should be an extended experience in which the applicant must develop skills in the specific types of test directions, questions and content found in formal testing.

Milhalka (1970) reported that the Ohio Bureau of Employment Services developed a structured pretest orientation program which included group discussion on reasons for testing, kinds of tests, scoring procedures, use of separate answer sheets, and test-taking hints. Short simple exercises provided practice with a variety of common tests found in industry, government, and education.

Finally, Erickson (1972) from careful observation as a reading teacher suggested techniques and procedures appropriate for providing students with test sophistication in answering control thought, detail, context, and inference type reading questions. He makes the suggestion that test sophistication skills need to be taught as an integral part of developmental, corrective, and remedial reading programs.

Implementing a Test-Wiseness Program into the Schools

Tests have been used increasingly in recent years by government, industry, and educational institutions in making selection, placement, and guidance decisions. But the increasing use of tests has been accompanied by an increasing flow of criticism directed toward test producers and consumers. The fairness of tests to underprivileged and minority groups, the heavy reliance on tests in making decisions, and the rigid and inflexible classification all too often placed on an individual are three of the most frequently expressed concerns stemming from the testing revolution.

There is no doubt that much can and should be done by test producers to improve their instruments, and by consumers of tests to accurately interpret and use test results. More importantly, however, is the need for testing services to de-emphasize the use of norms, percentile, etc., and for counselors, teachers, employers, and laymen to get a clearer understanding of what a person's test score represents. It should not be enough to allow an individual's test score to merely be viewed in terms of norms, scales, and other technical considerations, which in fact are only a small part of what a test score should reflect. As Dyer (1971) once observed:

"We have been so concerned with using test scores to classify students in one way or another or to predict how they will make out, academically or otherwise, that we have, I think, paid far too little attention to the details of behavior that a score presumably summarizes."

An individual's test score is a composite measure reflecting both ability and experience and as such is influenced by such environmental factors as the educational process. When combined with information about an individual's experiential background, test scores could facilitate a wide array of educational experiences for maximizing individual development.

From the topic developed in the preceding pages, test wiseness appears to be an important component of an individual's test score and seems amenable to the type of instructional program observed in specific courses in the schools. It follows, therefore, that (1) an attempt should be made to identify students who are low in test-wiseness, and (2) such students should then be instructed in the aspects of test wiseness previously described. The objective of such a learning program would be to decrease (1) some of the error of measurement resulting from the influence of test wiseness on test scores and (2) the handicap under which many examinees apparently operate.

Once a learning program in test wiseness has been developed, periodic practice and/or instruction in standardized and teacher made test-taking techniques for all grades would lend more credence, both educational and moral, to the practice of making decisions regarding an individual according to scores on group administered tests.

The questions now raised may be (1) from whom should the learning program be designed, (2) what format should it take, (3) who should be responsible for implementing the program, and (4) what might be the expected outcome.

Target Audience

In that the most widely used ETS tests are administered during the secondary school years and beyond, it seems reasonable to expect that a learning program

developed by ETS should be geared toward students at that level. It should not be considered, however, that younger children will not benefit as well from such an effort. The various strategies, principles, and related topics discussed throughout the paper appear to be applicable to a variety of testing situations (e.g. strategies for multiple-choice and analogy items, advisability of changing answers, etc.) and would seem to be adaptable to several different tests, if necessary.

Suggested Format

Although most of the test-wiseness instructional programs cited in the literature have taken the form of linear and branching programmed texts, this medium need not be the only approach that could be developed with ETS resources. However, since these texts are available, appropriate use can and should be made of them.

In an earlier effort, CEEB published a SAT Preparation Course with audio-cassette. It is being suggested that the various aspects of test-wiseness, outlined in this paper, be discussed extensively and programmatically on cassettes. Rather than present the content of the tape as a formal lecture, thought should be given to presenting the material in a manner interesting to young people (perhaps a hip dialogue between two persons) so as to maintain the attention of secondary school age audiences. Illustrative work materials should accompany the tape and thereby actively engage students in the learning program (For example, the tape recording could present the test-wiseness principle "always guess even if the usual correction or a more severe penalty for guessing is employed, whenever elimination of options provides sufficient chance of profiting"). The accompanying materials would present items such that students were required to learn to discriminate when (options can be eliminated) and when not (all options are plausible) to guess.

Video tapes are another form in which the test-wiseness learning program can be presented. Once the content of the program has been developed, a video tape of a teacher in a classroom presenting several lessons in test-wiseness could be made. This tape would be useful to students who wish to "boost" their test-wiseness score, or to pre-professional or in-service teachers to acquaint them to the topic of test-wiseness.

Kodak slides which capture the various aspects and principles of test-wiseness are additional devices that could be useful to teachers in classroom settings or to school counselors. The slides could flash on the screen the principles (e.g. eliminate those options which include specific determiners) and then several examples of this principle, in operation, could be presented to the students. The student might then be asked to write an item he felt would illustrate this principle.

Finally student handbooks could be distributed to a class of learners. Both the student handbook, and a guidebook to be used by classroom teachers or guidance counselors, might be developed to supplement each other, much as regular teacher and students texts are designed.

In addition to presenting the principles and strategies previously outlined on pages 9 and 10 empirical evidence and other programs in test wiseness that have been implemented, support the following additional topics which might be incorporated in a student handbook.

Implementation

As might have been implied by the preceding section, a test wiseness program could be designed to be implemented (1) by teachers who wish to incorporate

instruction in test wiseness into their regular classroom curriculum (2) by counselors in individual or small group guidance sessions prior to a major exam, (3) by individual students who wish to become more familiar with the art of taking tests, or (4) any combination of the above.

Suggested Content for Student Handbook

Based upon empirical research, the following items should be incorporated in a student handbook.

1. Practice tasks intended to permit familiarization with items under a variety of conditions:
 - a. sentence completion
 - b. antonyms
 - c. analogies
 - d. reading comprehension
2. Information as to why the test is administered.
3. Factors involved in scoring tests.
4. Explanation of test norms and percentile ranking (illustrate to students that with as few as three or four additional correct responses, one's percentile rank may increase as much as 10 points or more).
5. Discussion of what the test results will be used for.
6. Rationale of scoring formulas.
7. Examples of simulated directions similar to those the applicant will face in formal testing.
 - a. instruction in comprehending various oral and written directions.
8. Instruction in marking answers on separate answer sheets.
9. Simulated tests (discontinued test forms) in which the candidate becomes aware of working against time.

BIBLIOGRAPHY

- Anastasi, A., Psychological Testing. New York: The Macmillan Company, 1968.
- Bath, J., Answer - Changing Behavior on Objective Examinations, Journal of Educational Research, 1967, 61, 105-107.
- Berrein, F., Are First Impressions Best on Objective Tests?, School and Society, 1939, 50, 319-320.
- Callenbach, C., The Effects of Instruction and Practice in Content-Independent Test-Taking Techniques Upon the Standardized Reading Scores of Selected Second Grade Students, Journal of Educational Measurement, 1973, 10, 25-29.
- Coffman, W. and Parry, M., Effects of an Accelerated Reading Course on SAT-V Scores, Personnel and Guidance Journal, 1967, 46, 292-296.
- Coladarci, A., An Analysis of Miller Analogy Test Scores Changes, Educational and Psychological Measurement, 1960, 20, 817-823.
- College Entrance Examination Board., Effects of Coaching on Scholastic Aptitude Test Scores, New York: CEEB, 1965.
- Copeland, D., Should Chemistry Students Change Answers on Multiple-Choice Tests?, Journal of Chemical Education, 1972, 19, 258.
- Droege, R., Effects of Practice on Aptitude Scores, Journal of Applied Psychology, 1966, 50, 306-310.
- Dunn, T. and Goldstein, L., Test Difficulty, Validity, and Reliability as Functions of Selected Multiple-Choice Item Construction Principles, Educational and Psychological Measurement, 1959, 19, 171-179.
- Dyer, H., Does Coaching Help?, College Board Review, 1953, No. 19, 331-335.
- Dyer, H., Needed Changes to Sweeten the Impact of Testing, Personnel and Guidance Journal, 1967, 45, 776-780.
- Fbel, R., Measuring Educational Achievement, New Jersey: Prentice-Hall, 1965.
- Erickson, M., Test Sophistication: An Important Consideration, Journal of Reading, 1972, 16, 140-144.
- Flaughner, R., Research on Testing and the Minority Student. Paper presented at the GRE Board Research Seminar, 12th Annual Meeting of the Council of Graduate Schools, 1972, New Orleans, Louisiana.
- Frankel, E., Effects of Growth, Practice, and Coaching on Scholastic Aptitude Test Scores, Personnel and Guidance Journal, 1960, 38, 713-719.
- French, J., Effect of Anxiety on Verbal and Mathematical Exam Scores, Educational and Psychological Measurement, 1962, 22, 553-564.
- French, J. and Dear, R., Effect of Coaching on an Aptitude Test, Educational and Psychological Measurement, 1959, 19, 319-330.
- Gibb, B., Test-Wiseness as Secondary Cue Response, (Doctoral dissertation, Stanford University) 1964.
- Hammerton, M., The Guessing Correction in Vocabulary Test, British Journal of Educational Psychology, 1965, 35, 249-251.
- Kreit, L., The Effects of Test-Taking Practice on Pupil Test Performance, (Doctoral Dissertation, Indiana University, 1967).
- Lynch, D. and Smith, B., To Change or Not to Change Item Responses When Taking Tests; Empirical Evidence for Test Takers, Paper presented American Educational Research Association, Chicago, April, 1972.
- Lowe, M. and Crawford, C., First Impression vs. Second Thought in True-False Tests, Journal of Educational Psychology, 1929, 20, 192-195.
- Michael, E., Test Sophistication: An Important Consideration, Journal of Reading, 1972, 16, 140-144.

- Millman, J., Bishop, C., and Ebel, R., An Analysis of Test-Wiseness, Educational and Psychological Measurement, 1965, 25, 707-726.
- Millman, J. and Setijadi, A Comparison of the Performance of American and Indonesian Students on Three Types of Test Items, Journal of Educational Research, 1966, 59, 273-275.
- Milhalka, J., Pretesting Orientation for the Disadvantaged in Focus on the Future, American Personnel and Guidance Association, Washington, D.C., 1970, 273-274.
- Moore, J., Manipulating the Effectiveness of a Self-Instructional Program, Journal of Educational Psychology, 1968, 59, 315-319.
- Moore, J., Test Wiseness and Analogy Test Performance, Measurement and Evaluation in Guidance, 1971, 3, 198-202.
- Moore, J., Schultz, R., and Baker, R., The Application of a Self-Instructional Technique to Develop a Test-Taking Technique, American Educational Research Journal, 1966, 3, 13-17.
- Oakland, T., Effects of Test-Wiseness Materials on Standardized Test Performance of Pre-School Disadvantaged Children, Journal of School Psychology, 1972, 10, 355-360.
- Oakland, T., and Weilert, E., The Effects of Test-Wiseness Materials on Standardized Test Performance of Pre-School Disadvantaged Children, Paper presented at February 1971 Convention of the American Educational Research Association, New York.
- Omvig, C., Effects of Guidance on the Results of Standardized Achievement Testing, Measurement and Evaluation in Guidance, 1971, 4, 47-52.
- Preston, R., Ability of Students to Identify Correct Responses Before Reading, Journal of Educational Research, 1964, 58, 181-183.
- Reile, P. and Briggs, L., Should Students Change Their Initial Answers on Objective Tests? More Evidence Regarding an Old Problem, Journal of Educational Psychology, 1952, 43, 110-115.
- Roberts, S., and Oppenheim, D., The Effect of Special Instruction Upon Test Performance of High School Students in Tennessee. Report from CEEB Research and Development Division, 1966.
- Seiler, J., Preparing the Disadvantaged for Tests, Vocational Guidance Quarterly, 1971, 19, 201-205.
- Sheriffs, A., and Boomer, D., Who Is Penalized by the Penalty for Guessing?, Journal of Educational Psychology, 1954, 45, 81-90.
- Slakter, M., Effect of Guessing Strategy on Objective Test Scores, Journal of Educational Measurement, 1968, 5, 217-221.
- Slakter, M., Koehler, R., and Hampton, S., Grade Level, Sex, and Selected Aspects of Test-Wiseness, Journal of Educational Measurement, 1970a, 7, 119-122.
- Slakter, M., Koehler, R., and Hampton, S., Learning Test-Wiseness by Programmed Texts, Journal of Educational Measurement, 1970b, 7, 247-254.
- Stanley J., Reliability. in R. Thorndike, Educational Measurement, Washington, D.C., American Council on Education, 1971, 356-442.
- Taylor, P., A Study of the Effects of Instructions in a Multiple-Choice Mathematics Test, British Journal of Educational Psychology, 1966, 36, 1-6.
- Vernon, P., The Determinants of Reading Comprehension, Educational and Psychological Measurement, 1962, 22, 269-286.
- Wahlstrom, M., and Boersma, F., The Influence of Test-Wiseness Upon Achievement, Educational and Psychological Measurement, 1968, 28, 413-420.

Whitla, D., Effect of Tutoring on SAT Scores, Personnel and Guidance Journal, 1962, 41, 32-37.

Woodley, K., Test-Wiseness Program Development and Evaluation. Paper presented at the March 1973 Convention of the American Educational Research Association, New Orleans.